EFSO and DFS diagnostics for JMA’s global Data Assimilation System: their caveats and potential pitfalls

Daisuke Hotta and Yoichiro Ota

Numerical Prediction Division, Japan Meteorological Agency

Special thanks to:

many colleagues at JMA

Eugenia Kalnay (UMD) and Takemasa Miyoshi (UMD/RIKEN) for general advice

Sid Boukabara and James Jung (JCSDA) for guidance on using S4

The goal of today's talk

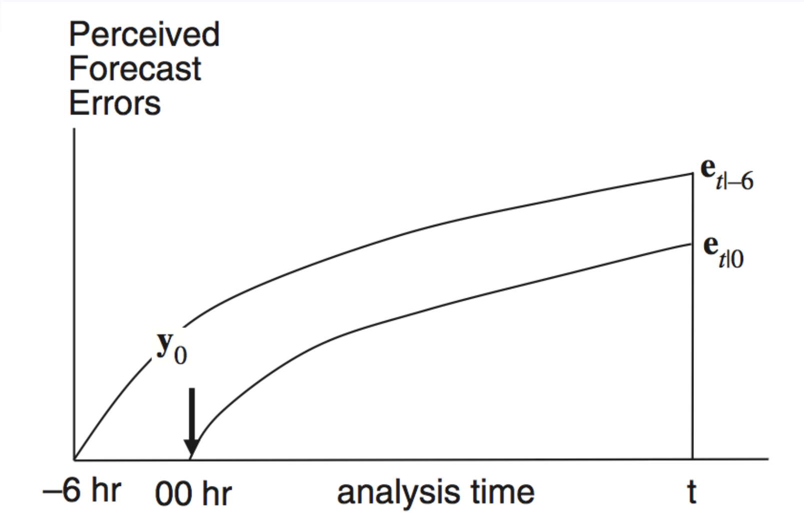
- To present recent progress on two diagnostics on Ensemble Data Assimilation: **EFSO** and **DFS**
- In particular,
- how we interpret their results
  - including potential pitfalls we have identified
- and what information we can get from the results to improve data assimilation system.

Part 1:

EFSO at JMA

1. Introduction: What is EFSO?

EFSO: Ensemble Forecast Sensitivity to Observations



- Langland and Baker (2004) introduced ***adjoint-based*** **FSO** method that enables to estimate how much each observation improved/degraded forecast without performing expensive data-denial experiments.
- Liu and Kalnay (2008) adapted FSO to LETKF.
- Kalnay et. al (2012) devised improved, simpler formulation applicable to ***any EnKF***.
- Ota et al. (2013) implemented the new EFSO into the NCEP's operational GFS system.

$$\Delta e^2 = \mathbf{e}_{t|0}^T C \mathbf{e}_{t|0} - \mathbf{e}_{t|-6}^T C \mathbf{e}_{t|-6} \approx \frac{1}{K-1} \delta \mathbf{y}_0^T \mathbf{R}^{-1} \mathbf{Y}_0^a \mathbf{X}_{t|0}^{fT} C(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6})$$

Reduction of forecast error by the assimilation of obs.

O-B of ens. mean

forecast ptbs.

analysis spread in obs. space

FSO and EFSO enables us to estimate how much *each* observation improved/degraded forecast

2. EFSO implementation at JMA

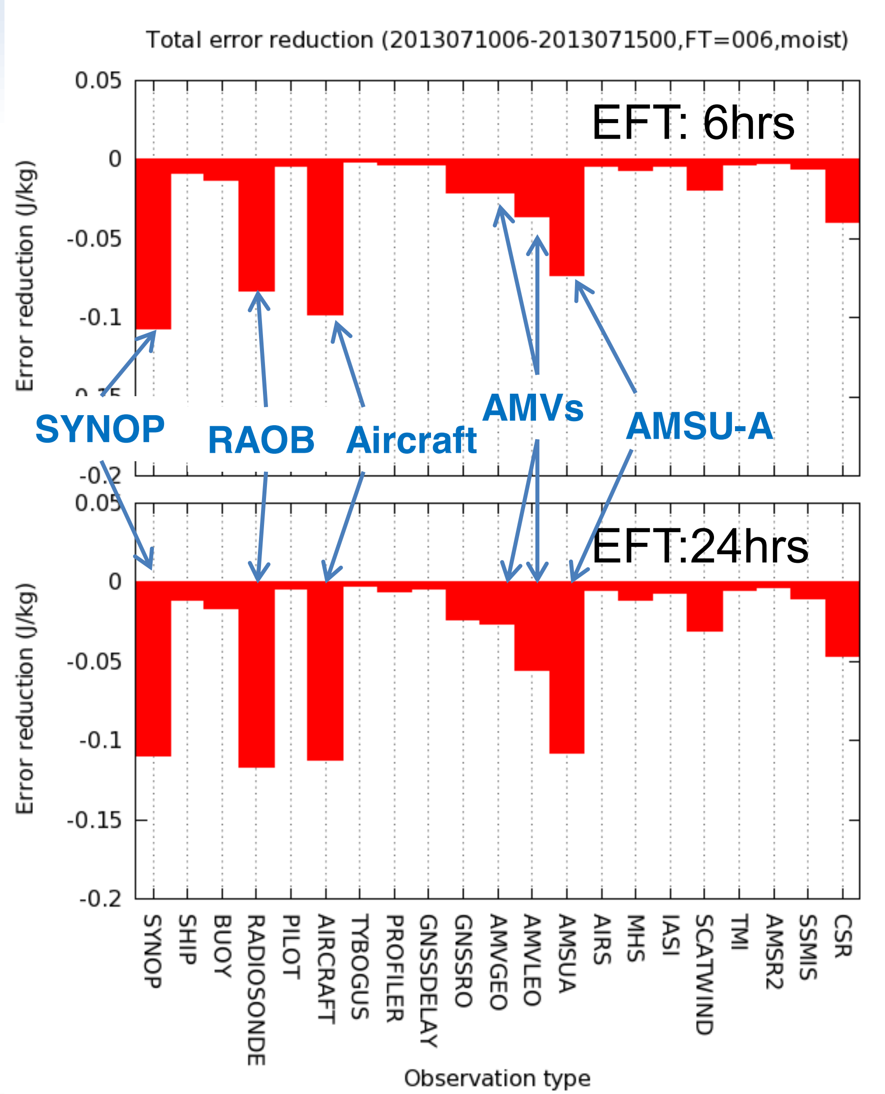
by Yoichiro Ota (2015)

- DA system: hybrid LETKF/4D-Var coupled with JMA GSM
  - Resolution: (outer) TL959L100 ; (inner and ensemble) T319L100
  - Window: 6 hours (analysis time +/- 3 hours)
  - **B** weights: 77% from static, 23% from ensemble
  - Member size: 50
  - Localization scales (e-folding):
    - LETKF: Horizontal: 400km, Vertical: 0.4 scale heights
    - 4D-Var: Horizontal: 800km, Vertical: 0.8 scale heights
  - Covariance Inflation: Adaptive inflation of Miyoshi (2011)
  - LETKF part initially coded by Dr. T. Miyoshi; maintained and updated by Y. Ota and T. Kadowaki.
- EFSO:
  - Lead-times investigated: FT=0,6,12,24
  - Localization scales: same as LEKTF
    - advection: “moving localization scheme” of Ota et al.(2013) with scaling factor of 0.5 for horizontal wind.
  - Verification: high-resolution analysis from 4D-Var
  - Error norm: KE, Dry TE and Moist TE
- Period: Jul. 10, 2013, 06UTC – Jul. 15, 2013, 18UTC (5days, 20cases)

3. Results

net EFSO contribution from each observation type (target=globe; norm=moist total energy)

- Overall, the results are consistent with other centers:
  - at FT=24, contributions from radiances and conventional data are comparable.
  - AMSU-A, Radiosonde SYNOP and Aircraft are the top contributors to fcst err reduction.
- Contributions from hyperspectral sounders (AIRS, IASI) are modest compared to ECMWF or NCEP.
- **6-hour EFSO is consistent with 24-hour EFSO.**



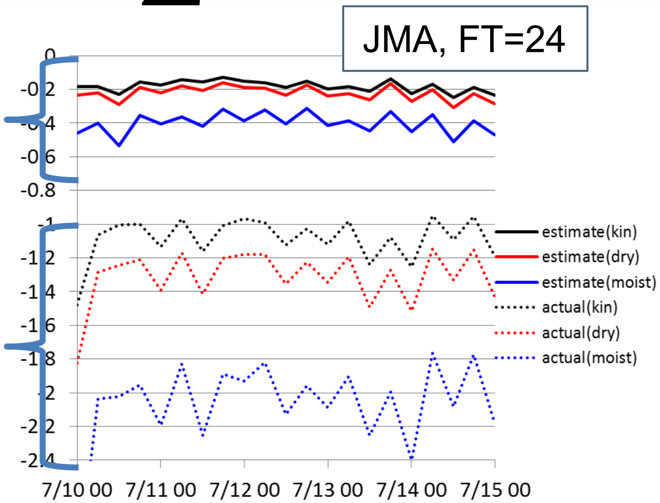
- So far, everything seems working well.
  - EFSO estimation consistent with other FSO studies
  - Plausible relative contributions from different observation types
- However,….

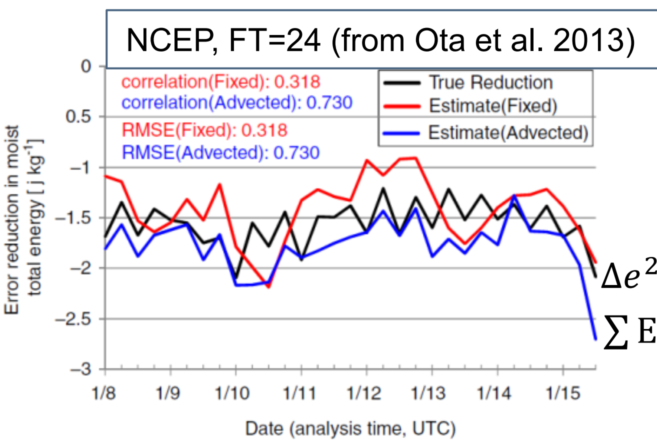
4. Underestimation problem: Estimated and actual forecast error reduction

$$\sum \mathrm{EFSO} \qquad \Delta e^2 = \frac{1}{2}\mathbf{e}_{t|0}^{f\mathrm{T}}\mathbf{C}\mathbf{e}_{t|0}^{f} - \frac{1}{2}\mathbf{e}_{t|-6}^{f\mathrm{T}}\mathbf{C}\mathbf{e}_{t|-6}^{f}$$

$\sum \mathrm{EFSO}$

Δe^2





Δe^2 (black)

$\sum \mathrm{EFSO}$ (blue and red)

- EFSO successfully reproduces temporal variation of forecast error reductions (correlation coefficient as high as ~ 0.8), but
- Only ~ 20 % of the amplitude explained by EFSO.
  - In contrast to > 100% (overestimation) for NCEP's pure EnKF (Ota et al. 2013)

5. A possible reason for impact underestimation (1/3)

- EFSO implemented for JMA's LETKF underestimates forecast error reduction, whereas, for NCEP's standalone EnKF, EFSO overestimates the actual impact.
- Why?
- Bug? → not found.
- Possible reason: **forecast error not well covered by the space spanned by the forecast perturbations**

5. A possible reason for impact underestimation (2/3)

- EFSO formulation: $\Delta e^{f-g} \approx \frac{1}{K-1}\mathbf{d}^T\mathbf{R}^{-1}\left[\rho \circ \mathbf{Y}^a\,\mathbf{X}^{f^T}\right]\mathbf{C}(\mathbf{e}^f_{t|0}+\mathbf{e}^f_{t|-6})$

- In evaluating

$$\mathbf{X}^{f^T}\mathbf{C}(\mathbf{e}^f_{t|0}+\mathbf{e}^f_{t|-6}) = \Big(\underbrace{\mathbf{C}^{1/2}\mathbf{X}^f}_{\widetilde{\mathbf{X}}^f}\Big)^T\underbrace{[\mathbf{C}^{\frac{1}{2}}(\mathbf{e}^f_{t|0}+\mathbf{e}^f_{t|-6})]}_{\tilde{\mathbf{e}}} =: \widetilde{\mathbf{X}}^{f^T}\tilde{\mathbf{e}}$$

  the portion of $\tilde{\mathbf{e}}$ that lies in the nullspace of $\widetilde{\mathbf{X}}^f$ does not contribute to $\widetilde{\mathbf{X}}^{f^T}\tilde{\mathbf{e}}$.
  Namely:

- Let $\tilde{\mathbf{e}} = \tilde{\mathbf{e}}_{\text{span}} + \tilde{\mathbf{e}}_{\text{null}}, \tilde{\mathbf{e}}_{\text{span}} \in \text{span}(\widetilde{\mathbf{X}}^f), \tilde{\mathbf{e}}_{\text{null}} \in \text{null}(\widetilde{\mathbf{X}}^f)$

  then

$$\widetilde{\mathbf{X}}^{f^T}\tilde{\mathbf{e}} = \widetilde{\mathbf{X}}^{f^T}(\tilde{\mathbf{e}}_{\text{span}} + \tilde{\mathbf{e}}_{\text{null}}) = \widetilde{\mathbf{X}}^{f^T}\tilde{\mathbf{e}}_{\text{span}}$$

- N.B.: This issue does not arise in adjoint FSO because, in its formulation

$$\Delta e^{f-g} \approx \mathbf{d}^T\mathbf{K}^T\mathbf{M}^T\mathbf{C}(\mathbf{e}^f_{t|0}+\mathbf{e}^f_{t|-6}) = \mathbf{d}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{P}^a\mathbf{M}^T\mathbf{C}(\mathbf{e}^f_{t|0}+\mathbf{e}^f_{t|-6})$$

  the matrix $\mathbf{P}^a\mathbf{M}^T$ is full rank.

5. A possible reason for impact underestimation (3/3)

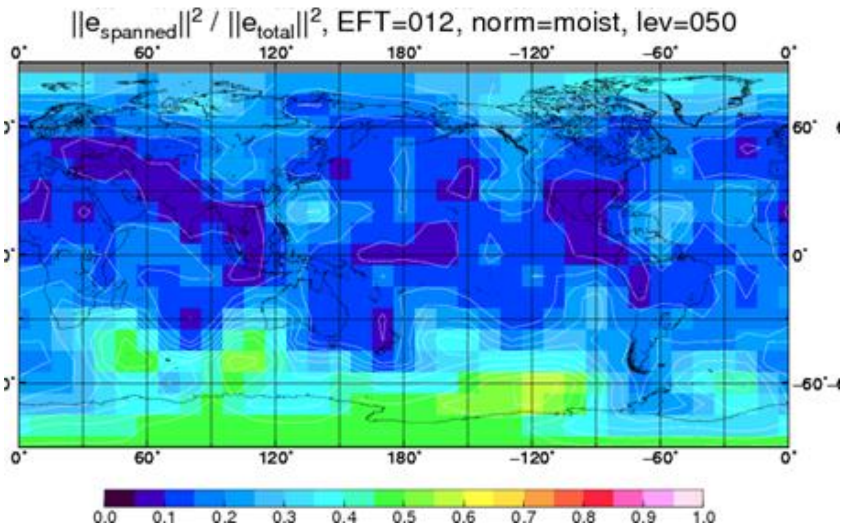
- Does this hypothesis really explain the impact underestimation?

→ Verify the hypothesis by performing the following diagnostics:

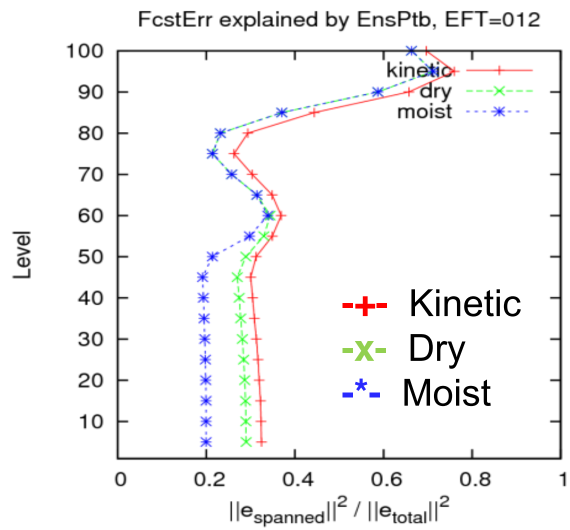
- For each local patch of LETKF,
  - Decompose $\tilde{\mathbf{e}}$ into $\tilde{\mathbf{e}}_{\mathrm{span}}$ and $\tilde{\mathbf{e}}_{\mathrm{null}}$. (detail in the backup slide)
  - Compute the "explained fraction" $\frac{\left\|\tilde{\mathbf{e}}_{\mathrm{span}}\right\|^2}{\|\tilde{\mathbf{e}}\|^2}$.
  - Compare this with the impact underestimation $\frac{\sum \mathrm{EFSO}}{\Delta e^2}$.
  - If the two agrees, we conclude that the hypothesis is likely correct.

Diagnosed "explained fraction" $\frac{\|\tilde{\mathbf{e}}_{\text{span}}\|^2}{\|\tilde{\mathbf{e}}\|^2}$

Horizontal distribution (near tropopause level)



Vertical Profile (global average)



- Fcst err well-captured by ensemble over the SH ocean, but not over the land.

→ Perhaps related to observation density:
  - Data-sparse area: analysis (verification) and forecast both close to model's free-run → $\mathbf{e}^f_{t|0}$ similar to Bred Vector → covered well by $\mathbf{X}^f$

- Errors in moisture difficult to capture by the ensemble.

Very good agreement between $\frac{\|\tilde{\mathbf{e}}_{\text{span}}\|^2}{\|\tilde{\mathbf{e}}\|^2}$ and $\frac{\sum \text{EFSO}}{\Delta e^2}$! (both ~ 20%)

6. New open questions

- Why only 20% of fcst err explained by ensemble in JMA's system?
  - tentative answers:
    - Not enough members (current 50 members perhaps insufficient)
    - **B**-weight assigned to ens too small (currently only 23%)
      → analysis increment largely aligned in $\mathrm{span}(\mathbf{B}_{\mathrm{static}})$
    - In fact, we found that for stand-alone EnKF, "explained fraction" $\|\tilde{\mathbf{e}}_{\mathrm{span}}\|^2/\|\tilde{\mathbf{e}}\|^2$ is ~ 70%.
- Does this underestimation mean that EFSO is less reliable than adjoint-based FSO?
  - No clear answer yet, but perhaps we should not be too pessimistic:
    - Fcst err $\tilde{\mathbf{e}}$ verified against analysis is different from true fcst err.
      - No difference in ens and adj in this respect.
    - Preliminary "apple-to-apple" comparison of EFSO and adj-FSO at JMA shows that they are similar in many aspects except their magnitude.
  - The "spanned-space" diagnostics presented here could be used to assess credibility of EFSO for each DA system.

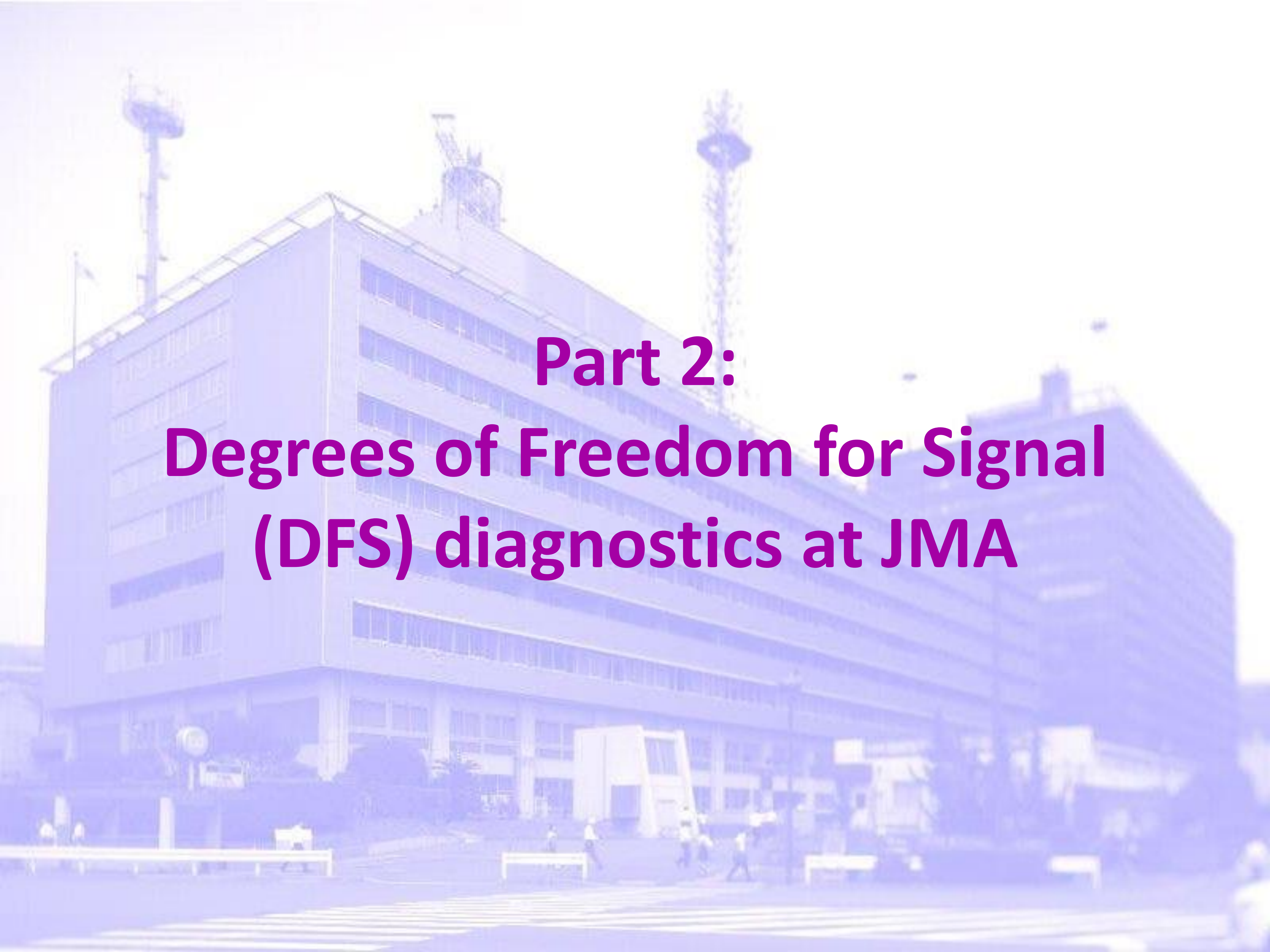
7. EFSO at JMA: Summary

- EFSO is successfully implemented on JMA's global DA system, both stand-alone LETKF and LETKF/4D-Var hybrid.
- Plausible impacts from different report types that are consistent with the literature.
- However, EFSO considerably underestimates the actual forecast error reduction Δe^{f-g}. **N.B: This is a system-specific problem. May not be applicable to other DAS (e.g. NCEP GFS)!**
- Diagnostics that decomposes the fcst err to column- and null-spaces of the fcst ensemble $\mathbf{X}^f$ suggests that the underestimation is caused because significant portion of fcst err lies in the null-space of $\mathbf{X}^f$.
- The diagnostics exposes the lack of the ensemble size (currently only 50).
  - → further corroborated by DFS diagnostics (Part 2)

8. EFSO at JMA: Future directions

- Better understand the nature of FSO and EFSO through
  - comparison with adjoint-based FSO implemented on JMA's 4D-Var
    - JMA is perhaps the only center which has both EFSO and adjoint FSO on the same DA system.
  - participation to **"FSOI Inter-comparison Project**" led by Dr. Auligné (JCSDA) and Dr. Gelaro (GMAO)
- Proactive QC
  - Identify detrimental observations using (E)FSO with short lead-time.
  - Repeat analysis without using them.
  - Proved successful using NCEP's EnKF-GSI hybrid 3D-Var on GFS.
  - c.f. Hotta (2014; UMD PhD dissertation), Hotta et al. (2016; in preparation for *MWR*)
- Tuning of **R** matrix using EFSR
  - EFSO can be extended to formulate forecast sensitivity to the **R** matrix.
  - This diagnostics gives guidance on how to optimize **R** via tuning.
  - Proved successful using NCEP's EnKF-GSI hybrid 3D-Var on GFS.
  - c.f. Hotta (2014; UMD PhD dissertation), Hotta et al. (2016; in preparation for *MWR*)
- Efficient design of observation pre-processing
  - Use (E)FSO instead of expensive OSE to deduce optimal QC criteria
  - Use (E)FSR to assign optimal **R** matrix
  - Proved effective with TMPA precipitation assimilation with LETKF (Lien et al. 2015ab, MWR)

  → Lien et al. (2016; in preparation for *MWR*)



Part 2:
Degrees of Freedom for Signal (DFS) diagnostics at JMA

1. Motivation

- How can we quantify the “value” of each observation?
- One possible quantification:
  - an observation is valuable if it improves the forecast.
- → FSO/EFSO

- Another approach (inspired from information theory):
  - An observation is valuable if it enhances our “knowledge” about the true state of the atmosphere.
  - Our “knowledge” is enhanced if the uncertainty of the state estimate is reduced by assimilating the observation.
  - → **Degrees of Freedom for Signal** (DFS, or information content).

2. What is DFS?

- Defined as the trace $\mathrm{tr}(\mathbf{S})$ of the "influence matrix" $\mathbf{S} = (\mathbf{HK})^{\mathbf{T}} = \frac{\partial \mathbf{y^a}}{\partial \mathbf{y^o}}$

- Shown to behave similarly to Shannon entropy reduction under some conditions (Fisher 2003, ECMWF tech memo #397):

$$\mathrm{tr}(\mathbf{S}) \approx \left[\, H(\mathbf{x}|\mathbf{x^b}) - H(\mathbf{x}|\mathbf{x^b}, \mathbf{y^o}) \,\right] \times \text{const.}$$

- Two ways to interpret:
    1. Analysis sensitivity to observations measured in obs space.
    2. The amount of information that the analysis extracted from observations.

Simple illustrative examples:

- **Forecast-Forecast cycle**: analysis is always the same as the background.
    - $\mathbf{y^a} \equiv \mathbf{y^b} \rightarrow \mathbf{S}$ is null, DFS=tr(S) = 0 (**0% information from obs.**)
- **Direct Insertion:** background is completely replaced by the obs.
    - $\mathbf{y^a} \equiv \mathbf{y^o} \rightarrow \mathbf{S}$ is identity, DFS = tr(S) = #obs
    - DFS per obs = 1 (**100% information comes from obs.**)

2. What is DFS?

- First introduced to NWP by Fisher (2003) and Cardinali et al. (2004)
- Popular diagnostics for variational DA systems.
  - Routinely monitored by several NWP centers (e.g. ECMWF, Météo-France)
- Liu et al. (2009) derived a simple method to compute DFS for EnKF:

$$\mathbf{S} = \frac{\partial \mathbf{y}^{\mathbf{a}}}{\partial \mathbf{y}^{\mathbf{o}}} = (\mathbf{HK})^{\mathbf{T}} = \mathbf{R}^{-1}\mathbf{HAH}^{\mathbf{T}} \approx \frac{1}{\mathbf{K}-1}\mathbf{R}^{-1}(\mathbf{Y}^{\mathbf{a}})(\mathbf{Y}^{\mathbf{a}})^{\mathbf{T}}$$

- Verified in Liu et al. (2009) with a simple AGCM (SPEEDY) in an idealized “identical-twin” scenario, but
- Up to present, not yet applied to operational Ensemble DA with real observations.

3. Ensemble-based DFS diagnostics at JMA
3-1. Experimental set-up

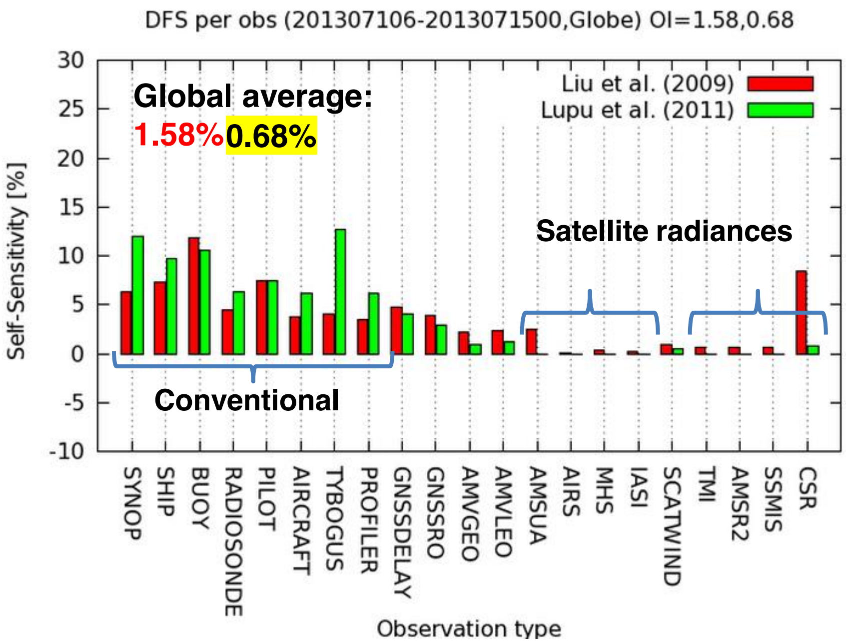
- DA system: hybrid LETKF/4D-Var coupled with JMA GSM
  - Resolution: (outer) TL959L100 ; (inner and ensemble) T319L100
  - Window: 6 hours (analysis time +/- 3 hours)
  - B weights: 77% from static, 23% from ensemble
  - Member size: 50
  - Localization scales (e-folding):
    - LETKF: Horizontal: 400km, Vertical: 0.4 scale heights
    - 4D-Var: Horizontal: 800km, Vertical: 0.8 scale heights
  - Covariance Inflation: Adaptive inflation of Miyoshi (2011)

identical to EFSO in Part 1

- DFS estimation Algorithms:
  - Liu et al. (2009) $\frac{\mathbf{1}}{\mathbf{K-1}}\mathrm{tr}\left(\mathbf{R}^{-\mathbf{1}}(\mathbf{Y}^{\mathbf{a}})(\mathbf{Y}^{\mathbf{a}})^{\mathbf{T}}\right)$
  - also tried the residual-based method of Lupu et al. (2011) as a double check:
    - $\mathrm{tr}(\mathbf{HK}) = \mathrm{tr}\left(\widetilde{\mathbf{R}}^{-1}\mathbb{E}\left(\mathbf{d}_{\mathrm{b}}^{\mathrm{a}}(\mathbf{d}_{\mathrm{a}}^{\mathrm{o}})^{\mathrm{T}}\right)\right), \widetilde{\mathbf{R}} = \mathbb{E}\left(\mathbf{d}_{\mathrm{a}}^{\mathrm{o}}(\mathbf{d}_{\mathrm{b}}^{\mathrm{o}})^{\mathrm{T}}\right)$ with the expectation evaluated as the average over a period and samples, assuming ergodicity and homogeneity

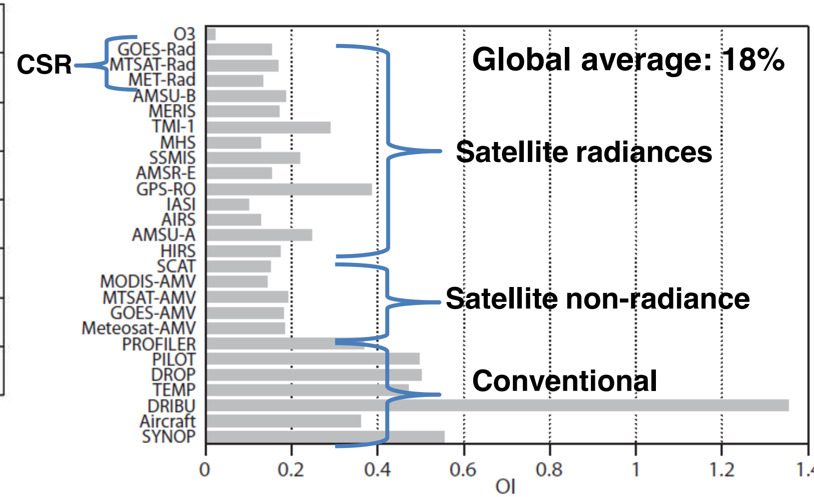
3. Ensemble-based DFS diagnostics at JMA
3-2. Results: **DFS per obs**

LETKF within JMA hybrid DA



c.f. ECMWF 4D-Var (as of 2011)

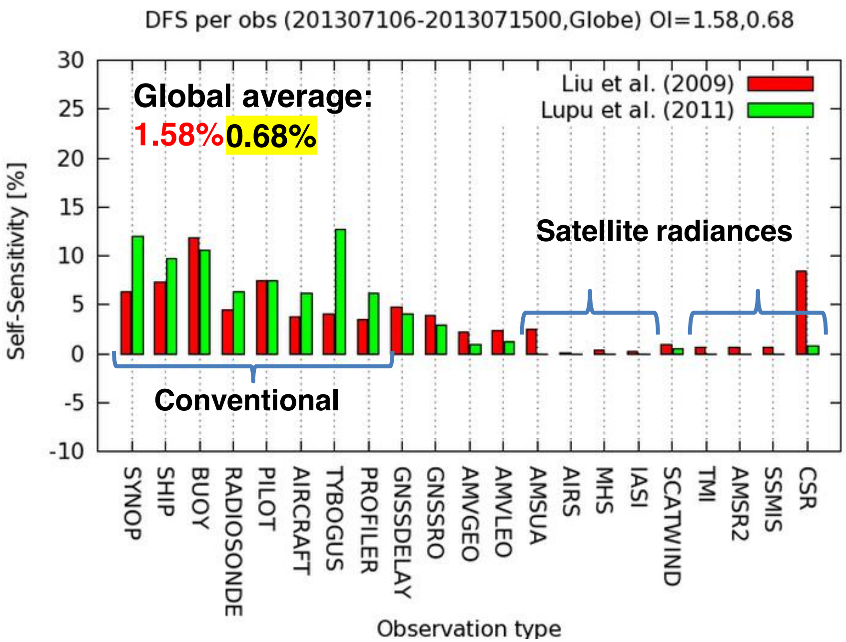
from Cardinali (2013; ECMWF lecture notes)



- Reasonable agreement between the two methods (at least for conventional obs).
- **Shockingly small observational impact**:
  - for JMA only about **1%** of information comes from observations,
  - whereas it is about 20% in ECMWF 4D-Var

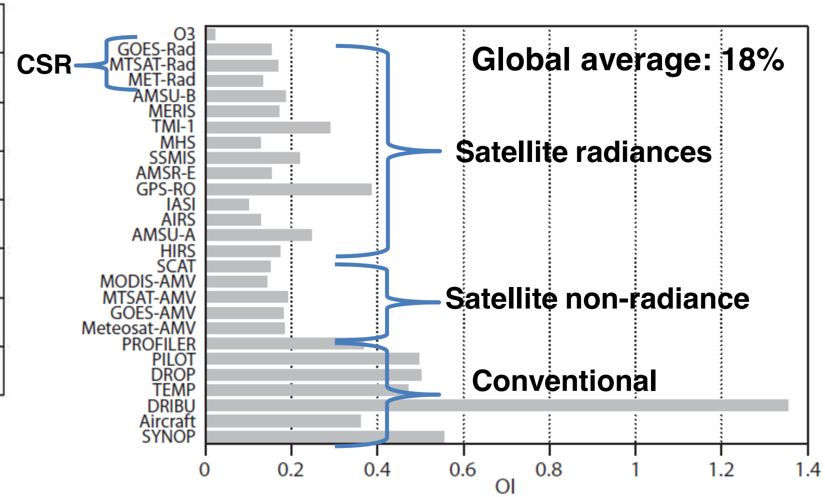
3. Ensemble-based DFS diagnostics at JMA
3-2. Results: **DFS per obs**

LETKF within JMA hybrid DA



c.f. ECMWF 4D-Var (as of 2011)

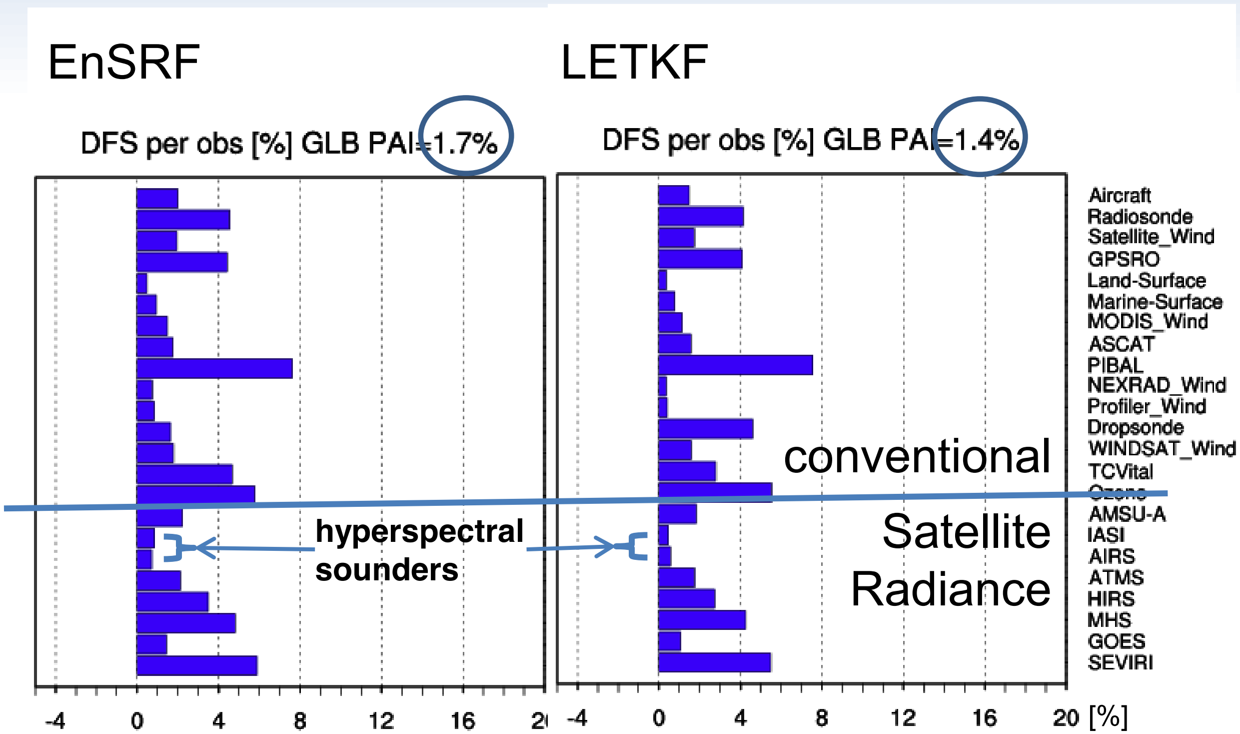
from Cardinali (2013; ECMWF lecture notes)



- DFS particularly small for dense observations, satellite radiances in particular (except AMSU-A and CSR*).

  * CSR: Clear Sky Radiances measured by infrared imagers on geostationary satellites (MTSAT, GOES and Meteosat)

4. Ensemble-based DFS for NCEP GFS hybrid GSI



- To discern if the "very small DFS problem" is merely an idiosyncrasy of JMA, we computed DFS for NCEP's lower-resolution version of GFS/GSI hybrid DA as well.
- Results: DFS is very small for NCEP's system as well.

→ **"Small DFS problem" possibly universal to all EnKF systems.**

5. Why DFS so small for EnKF?

- Our Answer: not enough ensemble size.
- We can show, for a local analysis in LETKF, that:

$$\mathrm{tr}(\mathbf{S}_{\mathrm{loc}}) = \mathrm{tr}\left(\mathbf{K}_{\mathrm{loc}}^{\mathrm{T}}\mathbf{H}_{\mathrm{loc}}^{\mathrm{T}}\right) = \mathrm{tr}(\mathbf{H}_{\mathrm{loc}}\mathbf{K}_{\mathrm{loc}}) \leq K - 1$$

- i.e., ***DFS is bounded from above by the degrees of freedom of the background ensemble***. See the next slide for proof.
- The number of observations locally assimilated, p_{loc}, is $\sim O(10^3)$, much larger than the member size $K = 50$.
- Suppose, for convenience, that each observation locally assimilated has comparable DFS, and that the observation density can be assumed homogeneous.
- Then, we can assume that, locally, (DFS per obs) $\sim \frac{K-1}{p_{\mathrm{loc}}}$, which gives:

$$\mathrm{DFS}_{\mathrm{global}} = \sum_{\mathrm{all\ obs}}(\mathrm{DFS\ per\ obs}\ )_{\mathrm{local}} \sim p_{\mathrm{global}} \times \frac{K-1}{p_{\mathrm{loc}}}$$

$\rightarrow$ $(\mathrm{DFS\ per\ obs}\ )_{\mathrm{global}} = \frac{\mathrm{DFS}_{\mathrm{global}}}{p_{\mathrm{global}}} \sim \frac{K-1}{p_{\mathrm{loc}}}$, which, for our system, is $\frac{49}{4{,}000} \sim O(1\%)$

6. Implications

- We have seen that, for an EnKF with ensemble size K much smaller than the number of the locally assimilated observations p_{loc} ($p_{\mathrm{loc}} \gg K$), DFS is inevitably bounded by the member size K and hence automatically underestimated.
- This means that such a system cannot fully extract information from observations.
- We believe this fact has a lot of important implications, e.g., on:
    1. why drastic ***observation thinning does not harm performance***,
    2. why ***covariance inflation*** is necessary,
    3. ***what the localization scale should be***, given the ensemble size and observation density,
    4. how, ***in serial assimilation,*** the ***order*** of assimilating observations affects the accuracy of the analysis …etc.

6-1. Implication for observation thinning *(highly speculative)*

- Hamrud et al. (2015 Part I; MWR) reports that, **in ECMWF's LETKF local analysis, limiting the number of assimilated observations to only 30 per report type and element does not harm (*even improves*) forecast performance** while achieving dramatic computational saving at the same time.
  - Similar result was also obtained with JMA's LETKF (Ota 2015, "adjoint Workshop").
- This fact "using less obs is better" seems counterintuitive and difficult to interpret (at least to me).
- DFS discussion could provide a plausible interpretation (justification):
  - In LETKF local analysis, the amount of information extractable from observations (=DFS) is limited by the ensemble member size.
  - Thus, assimilating too many observations beyond this limit only adds noises rather than signal.
  - → Assimilating observations within the limit of DFS imposed by the member size reduces noises and improves analysis.
- Related to the argument above, in a situation where thinning of observations is necessary (e.g., very dense observation such as satellite hyperspectral sounding, radar data, etc.), it would make sense to assimilate only the observations with large DFS.

6-2. Implication on covariance inflation *(highly speculative)*

- If the ensemble size is insufficient, DFS=$\mathrm{tr}(\mathbf{R}^{-1}\mathbf{HAH}^{\mathrm{T}})$ is underestimated.
- → The analysis error covariance $\mathbf{A}$ is also underestimated.
- → Need to inflate $\mathbf{A}$.

- Traditionally, nonlinearity and model errors are considered to be the source of necessity for covariance inflation
  - It is **B** rather than **A** that need inflation.
  - This is true for Extended Kalman Filter.
- The inherent underestimation of DFS could be another mechanism behind the need for covariance inflation.
- This argument gives intuitive explanation as to why Relaxation-to-prior methods of Zhang et al. (2004) and Whitaker and Hamil (2012) are so successful:
  - underestimation of DFS (= ***posterior spread in obs space***) is severer when/where observations are denser
  - Relaxation-to-prior methods act to inflate **A** more exactly in such a situation.

6-3. Implication on covariance localization *(highly speculative)*

- Traditionally, it is believed that localization is necessary to filter out spurious correlations in **B** due to sampling errors.
- From this perspective, observation density/distribution does not come into play.

- The fact that DFS is bounded by the member size provides another criterion for optimality of localization:
  - Let $\{\sigma_i\}$ be the singular values of $\mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{B}^{\frac{1}{2}}\left(=\frac{\mathbf{1}}{\sqrt{\mathrm{K}-1}}\mathbf{R}^{-\frac{1}{2}}\mathbf{Y}^{\mathbf{b}}\right)$. Then, DFS = $\sum_i \frac{\sigma_i^2}{1+\sigma_i^2}$
  - → DFS will not be underestimated if K-th largest singular value σ_K is negligibly small.
- This gives a criterion for the optimal member size K given the observation network $(\mathbf{H}, \mathbf{R})$ and background error covariance $(\mathbf{B})$.
- Inversely, given the member size K, we can choose localization scale so that DFS is not artificially bounded. For this, we can require that the observations within the localized area are few enough such that $\sigma_i \ll 1$ for some $i < K$.

6-4. Implication on order of obs. assimilation in serial EnKF *(highly speculative)*

- Given that the total DFS is bounded by the ensemble size, it would make sense to assimilate the observation with the largest DFS first.
- In fact, Dr. Jeff Whitaker showed at ISDA 2015 that, in serial EnSRF, the following procedure improves the analysis:
  - assimilating observations from those with the smallest $\rho := \frac{\mathrm{HAH}^T}{\mathrm{HBH}^T}$ to those with the largest,
  - assigning large localization scale to observations whos ρ is small.
- It is easy to see $\rho = \frac{\mathrm{HAH}^T}{\mathrm{HBH}^T} = \frac{\mathrm{H(I{-}KH)BH}^T}{\mathrm{HBH}^T} = 1 - \mathrm{HK}$ =1-DFS, i.e., Dr. Whitaker's successful method is equivalent to:
  - assimilating observations from largest DFS to those with smallest,
  - assigning a larger localization scale to observations with larger DFS
- → DFS argument could provide a theoretical justification to his somewhat empirical but successful method.

7. Summary of DFS diagnostics

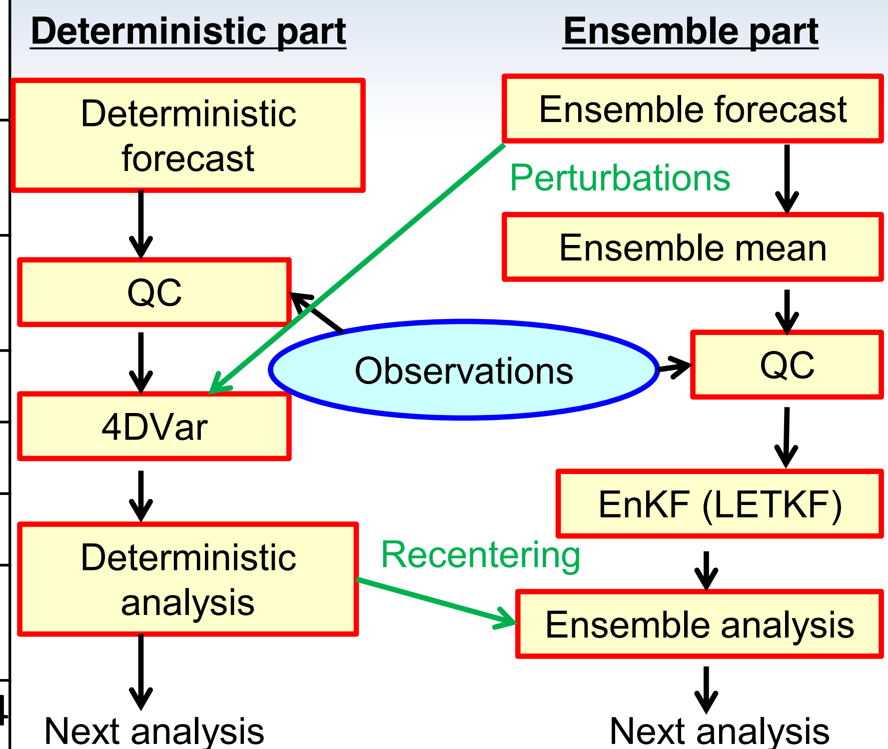
- Ensemble based DFS estimation of Liu et al. (2009) is implemented for the first time (perhaps) to a quasi-operational DA system with real data.
- In order for comparison, DFS was also computed using the residual-based method of Lupu et al. (2011).
- DFS computed with the both methods turned out to be very small compared to that of variational method.
- Simple mathematical manipulation shows that DFS is automatically underestimated (bounded from above by the member size) if the number of locally-assimilated observations is much larger than the member size ($p_{\mathrm{loc}} \gg K$).
- Underestimation of DFS is, by definition, directly related to underestimation of analysis spread.
- → This Entails implications to many important issues of EnKF including covariance inflation/localization and observation thinning.

Backup slides

from Yoichiro Ota (2015, Adjoint Workshop)

Hybrid 4DVar-LETKF DA developed in JMA

Analysis resolution (outer / inner)	T_L959L100 (~20km, top:0.01hPa) / T_L319L100 (~55km, top:0.01hPa)
Assimilation window	6 hours (analysis time +/- 3 hours)
Hybrid method	Extended control variable method (Lorenc 2003)
Weights on $\mathbf{B}$	$\beta_{stat}{}^2 = 0.77$, $\beta_{ens}{}^2 = 0.23$
LETKF resolution	T_L319L100
Ensemble size	50
Localization scale (4DVar)	Horizontal: 800km Vertical: 0.8 scale heights
Localization scale (LETKF)	Horizontal: 400km, Vertical: 0.4 (0.8 for Ps) scale heights
Covariance inflation	Adaptive inflation (Miyoshi 2011)



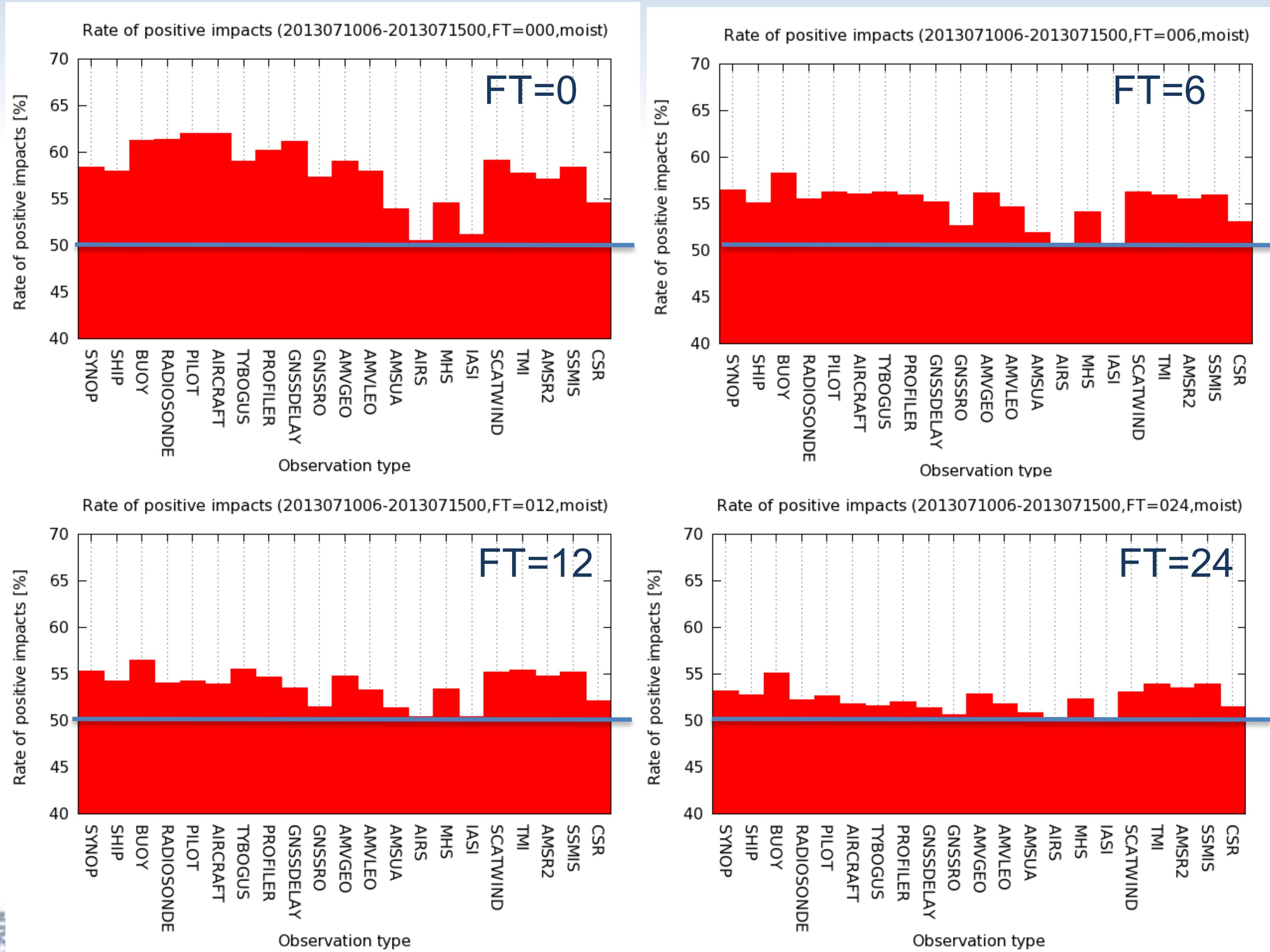
$$\mathbf{B} = \beta_{stat}^2 \mathbf{B}_{stat} + \beta_{ens}^2 \mathbf{B}_{ens}$$

Static (Climatological) background error covariance

Ensemble-based background error covariance

Operational global DA at JMA is 4DVar (not hybrid)

Percentage of positively-contributing observations (target=globe; norm=moist total energy)



- At FT=24, only slightly more than 50% of the observations contribute to improve forecast (as pointed out by many FSO studies in the literature).
- Percentage of “helpful” observations increases for shorter evaluation lead-time.
- → Consistent with Hotta (2014, PhD. Dissertation).

Decomposition of fcst error into column- and null- spaces of fcst ptbs

- Fix a grid and consider a local patch that would be used if an observation was located at the grid point in question. In the derivation below, all vectors/matrices are assumed to be restricted to this local patch.
- In EnKF, the sum of each column of $\mathbf{X}^f$ is zero, so rank($\mathbf{X}^f$)=$K-1$:

  $$\mathrm{span}(\widetilde{\mathbf{X}}^f) = \mathrm{span}\left([\widetilde{\mathbf{X}}^f]_1, \cdots, [\widetilde{\mathbf{X}}^f]_{K-1}, [\widetilde{\mathbf{X}}^f]_K\right) = \mathrm{span}\left([\widetilde{\mathbf{X}}^f]_1, \cdots, [\widetilde{\mathbf{X}}^f]_{K-1}\right)$$

In light of this, we now denote by $\widetilde{\mathbf{X}}^f$ the first $K-1$ columns of the original $\widetilde{\mathbf{X}}^f$.

- Now, suppose that $\widetilde{\mathbf{e}} := \mathbf{C}^{\frac{1}{2}}(\mathbf{e}^f_{t|0}+\mathbf{e}^f_{t|-6})$ can be decomposed as
- $\widetilde{\mathbf{e}} = \widetilde{\mathbf{e}}_{\mathrm{span}} + \widetilde{\mathbf{e}}_{\mathrm{null}}$, $\widetilde{\mathbf{e}}_{\mathrm{span}} = \sum_{\boldsymbol{k}=\mathbf{1}}^{\boldsymbol{K}-\mathbf{1}} \alpha_k [\widetilde{\mathbf{X}}^f]_{\boldsymbol{k}} = \widetilde{\mathbf{X}}^f \boldsymbol{\alpha}$,

  $\boldsymbol{\alpha} = (\alpha_1, \quad \cdots \quad \alpha_{K-1})^T$

  Multiplying $\widetilde{\mathbf{e}} = \widetilde{\mathbf{e}}_{\mathrm{span}} + \widetilde{\mathbf{e}}_{\mathrm{null}}$ with $\left(\mathbf{C}^{1/2}\mathbf{X}^f\right)^T =: \widetilde{\mathbf{X}}^{f\,T}$ from left, $\widetilde{\mathbf{e}}_{\mathrm{null}}$ vanishes by definition , giving:

  $$\widetilde{\mathbf{X}}^{f\,T}\widetilde{\mathbf{e}} = \widetilde{\mathbf{X}}^{f\,T}\left(\widetilde{\mathbf{X}}^f\boldsymbol{\alpha} + \widetilde{\mathbf{e}}_{\mathrm{null}}\right) = \widetilde{\mathbf{X}}^{f\,T}\widetilde{\mathbf{X}}^f\boldsymbol{\alpha}$$

  $$\therefore \boldsymbol{\alpha} = \left(\widetilde{\mathbf{X}}^{f\,T}\widetilde{\mathbf{X}}^f\right)^{-1}\widetilde{\mathbf{X}}^{f\,T}\widetilde{\mathbf{e}}$$

- Once $\boldsymbol{\alpha}$ is determined, we can obtain $\left\|\widetilde{\mathbf{e}}_{\mathrm{span}}\right\|^2$ and $\|\widetilde{\mathbf{e}}_{\mathrm{null}}\|^2$ by

  $$\left\|\widetilde{\mathbf{e}}_{\mathrm{span}}\right\|^2 = \left\|\widetilde{\mathbf{X}}^f\boldsymbol{\alpha}\right\|^2 = \left(\widetilde{\mathbf{X}}^f\boldsymbol{\alpha}\right)^T\left(\widetilde{\mathbf{X}}^f\boldsymbol{\alpha}\right) = \boldsymbol{\alpha}^T\widetilde{\mathbf{X}}^{f\,T}\widetilde{\mathbf{X}}^f\boldsymbol{\alpha} = \boldsymbol{\alpha}^T\widetilde{\mathbf{X}}^{f\,T}\widetilde{\mathbf{e}}$$

  $$\|\widetilde{\mathbf{e}}_{\mathrm{null}}\|^2 = \|\widetilde{\mathbf{e}}\|^2 - \left\|\widetilde{\mathbf{e}}_{\mathrm{span}}\right\|^2$$

Proof of $\mathrm{tr}(\mathbf{S}_{\mathrm{loc}}) \equiv \mathrm{tr}(\mathbf{H}_{\mathrm{loc}}\mathbf{K}_{\mathrm{loc}}) \leq K-1$ for LETKF local analysis

- In each local analysis of LETKF, DFS can be expressed as
- $\mathrm{tr}(\mathbf{S}) \equiv \mathrm{tr}(\mathbf{HK}) = \mathrm{tr}(\mathbf{HAH}^T\mathbf{R}^{-1})$ $(\because \mathbf{K} = \mathbf{AH}^T\mathbf{R}^{-1})$
- LETKF estimates the analysis error covariance by :

$$\mathbf{A} = \mathbf{X^b}\widetilde{\mathbf{A}}\mathbf{X^b}^T, \quad \widetilde{\mathbf{A}} = \left[(K-1)\mathbf{I} + \mathbf{Y^b}^T\mathbf{R}^{-1}\mathbf{Y^b}\right]^{-1} = \frac{\mathbf{1}}{K-1}(\mathbf{I} + \mathbf{Z}^T\mathbf{Z})^{-1}, \text{ with } \mathbf{Z} \equiv \frac{\mathbf{1}}{\sqrt{K-1}}\mathbf{R}^{-\frac{1}{2}}\mathbf{Y^b}$$

- $\mathbf{Z}^T\mathbf{Z}$ is a $K \times K$ positive semi-definite symmetric matrix. Its eigenvalue decomposition becomes :

$$\mathbf{Z}^T\mathbf{Z} = \mathbf{U\Lambda U}^{-1}, \mathbf{UU}^{-1} = \mathbf{I}, \qquad \mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_K)$$

- Since $\mathrm{rank}(\mathbf{Z}) = K-1$, $\lambda_K = 0$. From positive semi-definiteness of $\mathbf{Z}^T\mathbf{Z}$, $\lambda_i > 0$ $(1 \leq i \leq K-1)$.
- Thus:

$$\mathbf{HK} = \mathbf{HAH}^T\mathbf{R}^{-1} = \mathbf{HX^b}\widetilde{\mathbf{A}}\mathbf{X^b}^T\mathbf{H}^T\mathbf{R}^{-1} = \mathbf{Y^b}\widetilde{\mathbf{A}}\mathbf{Y^b}^T\mathbf{R}^{-1}$$

$$= \left(\sqrt{K-1}\mathbf{R}^{\frac{1}{2}}\mathbf{Z}\right)\widetilde{\mathbf{A}}\left(\sqrt{K-1}\mathbf{R}^{\frac{1}{2}}\mathbf{Z}\right)^T\mathbf{R}^{-1} = \mathbf{R}^{\frac{1}{2}}\mathbf{Z}(\mathbf{I} + \mathbf{Z}^T\mathbf{Z})^{-1}\left(\mathbf{R}^{\frac{1}{2}}\mathbf{Z}\right)^T\mathbf{R}^{-1}$$

- Because trace is invariant under cyclic reordering,

$$\mathrm{tr}(\mathbf{S}) \equiv \mathrm{tr}(\mathbf{HK}) = \mathrm{tr}\left(\mathbf{R}^{\frac{1}{2}}\mathbf{Z}(\mathbf{I} + \mathbf{Z}^T\mathbf{Z})^{-\mathbf{1}}\left(\mathbf{R}^{\frac{1}{2}}\mathbf{Z}\right)^T\mathbf{R}^{-1}\right) = \mathrm{tr}\left(\mathbf{Z}(\mathbf{I} + \mathbf{Z}^T\mathbf{Z})^{-\mathbf{1}}\mathbf{Z}^T\mathbf{R}^{\frac{1}{2}}\mathbf{R}^{-1}\mathbf{R}^{\frac{1}{2}}\right) =$$

$$\mathrm{tr}(\mathbf{Z}(\mathbf{I} + \mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T) = \mathrm{tr}\left((\mathbf{I} + \mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^\mathbf{T}\mathbf{Z}\right) = \mathrm{tr}\left(\left(\mathbf{I} + \mathbf{U\Lambda U}^{-\mathbf{1}}\right)^{-1}\mathbf{U\Lambda U}^{-1}\right) =$$

$$\sum_{i=1}^{K}\frac{\lambda_i}{1+\lambda_i} = \frac{\lambda_1}{1+\lambda_1} + \cdots + \frac{\lambda_{K-1}}{1+\lambda_{K-1}} + \frac{0}{1+0} \leq K-1$$